Deep Learning - Week 6

1. What is/are the primary advantages of Autoencoders over PCA?

   (a) Autoencoders are less prone to overfitting than PCA.

   (b) Autoencoders are faster and more efficient than PCA.

   (c) Autoencoders require fewer input data than PCA.

   (d) Autoencoders can capture nonlinear relationships in the input data.

   Correct Answer: (d)
   Solution: Autoencoders can capture nonlinear relationships in the input data, which allows them to learn more complex representations than PCA. This can be particularly useful in applications where the input data contains nonlinear relationships that cannot be captured by a linear method like PCA.

2. Which of the following is a potential advantage of using an overcomplete autoencoder?

   (a) Reduction of the risk of overfitting

   (b) Faster training time

   (c) Ability to learn more complex and nonlinear representations

   (d) To compress the input data

   Correct Answer: (c)

   Solution: Overcomplete autoencoders have more hidden units in the encoder than in the decoder, which can increase the capacity of the network and allow it to learn more complex and nonlinear representations of the input data.

3. We are given an autoencoder A. The average activation value of neurons in this network is 0.015. The given autoencoder is

   (a) Contractive autoencoder

   (b) Sparse autoencoder

   (c) Overcomplete neural network

   (d) Denoising autoencoder

   Correct Answer: (b)
   Solution: The neurons are mostly inactive for a given input. Hence the autoencoder is sparse autoencoder.

4. Suppose we build a neural network for a 5-class classification task. Suppose for a single training example, the true label is [0 1 0 0 1] while the predictions by the neural network are [0.4 0.25 0.2 0.1 0.6]. What would be the value of cross-entropy loss for this example? (Answer up to two decimal places, Use base 2 for log-related calculations)

Correct Answer: range(2.7, 2.8)
Solution: Cross entropy loss is given by $-\sum_{i=1}^{5}(y_i)\log_2(\hat{y}_i)$
$= -0 \cdot \log_2(0.4) - 1 \cdot \log_2(0.25) - 0 \cdot \log_2(0.2) - 0 \cdot \log_2(0.1) - 1 \cdot \log_2(0.6)$
$= -1 \cdot \log_2(0.25) - 1 \cdot \log_2(0.6)$
$= -1 \cdot -2 - 1 \cdot -0.7369$
$= 2.7369$

5. If an under-complete autoencoder has an input layer with a dimension of 5, what could be the possible dimension of the hidden layer?

   (a) 5

   (b) 4

   (c) 2

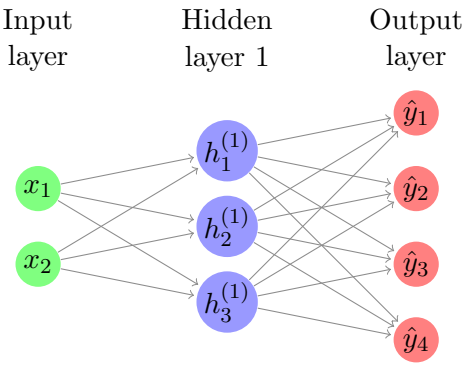
   (d) 0

   (e) 6

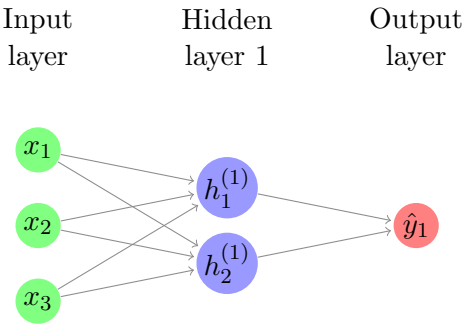
   Correct Answer: (b),(c)
   Solution: The dimension of the hidden layer is less than the input layer in the under-complete autoencoder.

6. Which of the following networks represents an autoencoder?

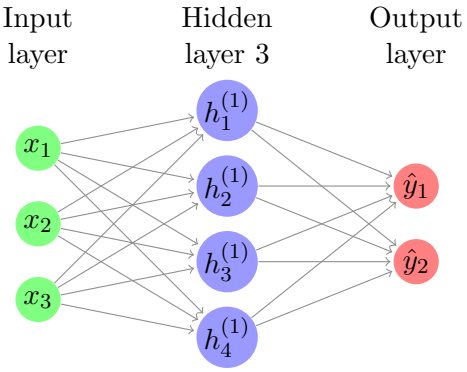
(a)



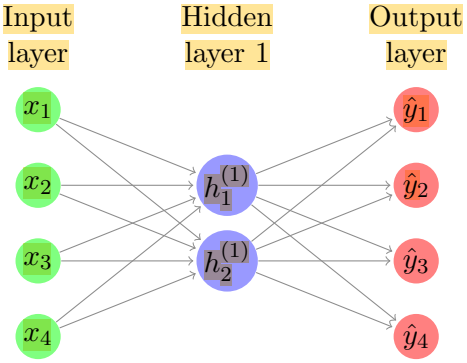
(b)



(c)



(d)



Correct Answer: (d)

Solution: Autoencoder is used to learn the representation of input data. Hence the output layer's size should be the same as the input layer's size to compare the reconstruction error.

7. What is the primary reason for adding corruption to the input data in a denoising autoencoder?

   (a) To increase the complexity of the model.

   (b) To improve the model's ability to generalize to unseen data.

   (c) To reduce the size of the training dataset.

   (d) To increase the training time.

Correct Answer: b)

Solution: Adding corruption to the input data in a denoising autoencoder serves the purpose of forcing the model to learn robust features that can reconstruct the original input even when parts of it are missing or noisy. This process prevents the model from merely memorizing the training data, thereby enhancing its ability to generalize to new, unseen data. This generalization is crucial for the model's performance on real-world tasks where the input may not always be clean or complete.

8. Suppose for one data point we have features x_1, x_2, x_3, x_4, x_5 as $-4, 6, 2.8, 0, 17.3$ then, which of the following function should we use on the output layer(decoder)?

   (a) Linear

(b) Logistic

(c) Relu

(d) Tanh

Correct Answer: (a)
Solution: The linear activation function is commonly used in regression tasks where the output can be any real number, which aligns with the nature of the given features. It allows the model to predict values across the entire real number line, which is suitable for the diverse range of input values we see in the features.

9. Which of the following statements about overfitting in overcomplete autoencoders is true?

(a) Reconstruction error is very high while training

(b) Reconstruction error is very low while training

(c) Network fails to learn good representations of input

(d) Network learns good representations of input

Correct Answer: (b),(c)
Solution: (b) Reconstruction error is very low while training: An overcomplete autoencoder, with more neurons in the hidden layer than in the input layer, has enough capacity to memorize the training data. This often results in a very low reconstruction error during training because the network effectively "copies" the input.

(c) Network fails to learn good representations of input: Although the autoencoder can reconstruct the training data accurately (due to memorization), it may not capture meaningful or generalizable features from the data. This means it fails to learn good representations that can be useful for tasks like dimensionality reduction or feature extraction on unseen data.

(a) is incorrect because the reconstruction error during training is low, not high.

(d) is incorrect because while the network learns to reconstruct inputs, it doesn't necessarily learn useful or robust representations.

10. What is the purpose of a decoder in an autoencoder?

(a) To reconstruct the input data

(b) To generate new data

(c) To compress the input data

(d) To extract features from the input data

Correct Answer: (a)
Solution: The decoder in an autoencoder is responsible for reconstructing the input data from the encoded representation generated by the encoder. It is used for data reconstruction and is typically the reverse of the encoding process.